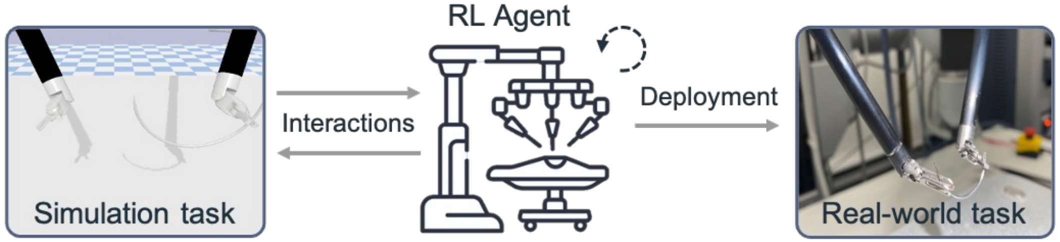
Demonstration-Guided Reinforcement Learning with Efficient Exploration for Task Automation of Surgical Robot

Tao Huang, Kai Chen, Bin Li, Yun-Hui Liu, Qi Dou

The Chinese University of Hong Kong

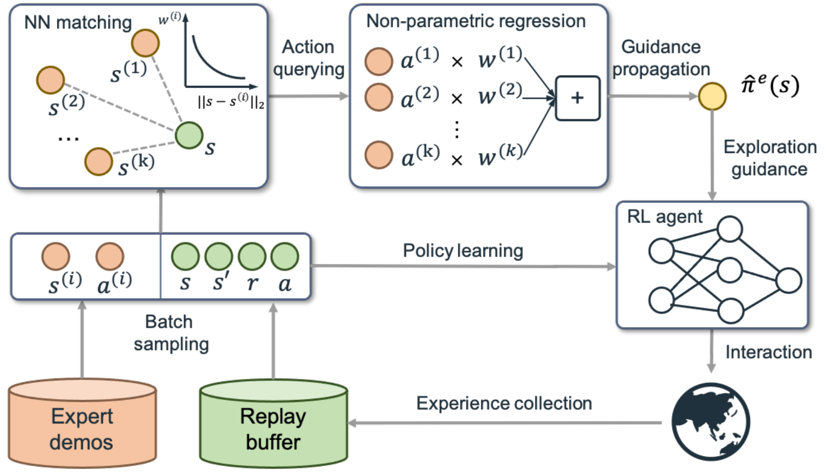
ICRA LONDON · 2023

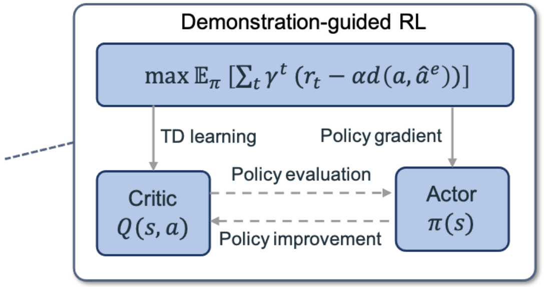
Background



- Reinforcement learning faces with exploration challenge in task automation of surgical robot
- Demonstration data can accelerate exploration, but how to make effective use of it remains a open problem

Method





- **Critic regularization** to facilitate exploration on expert-alike trajectories
- **Guidance propagation** to guide exploration at undemonstrated states

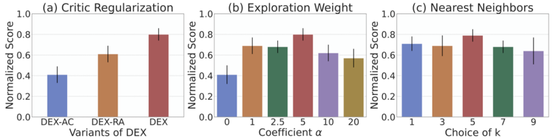
Main Results

	Task Description		Reinforcement Learning		Imitation Learning			Demonstration-guided Reinforcement Learning				Ours
	Task	$\mathcal{S}/\mathcal{A}/r$	SAC	DDPG	BC	SQIL	VINN	DDPGBC	AMP	CoL	AWAC	DEX
ECM	Aggregate	–	**0.99** (±.03)	**0.99** (±.02)	**1.00** (±.00)	0.24 (±0.06)	0.58 (±.06)	**1.00** (±.00)	**1.00** (±.01)	**1.00** (±.00)	**0.99** (±.01)	**1.00** (±.00)
	ECMReach	$\mathbb{R}^{12}/\mathbb{R}^3/S$	**1.00** (±.06)	**1.00** (±.00)	**1.00** (±.00)	0.07 (±.04)	0.49 (±.10)	**1.00** (±.00)	**0.99** (±.02)	**1.00** (±.00)	**1.00** (±.00)	**1.00** (±.00)
	StaticTrack	$\mathbb{R}^{16}/\mathbb{R}^3/S$	0.92 (±.14)	**0.98** (±.05)	**1.00** (±.00)	0.43 (±.26)	0.56 (±.10)	**1.00** (±.00)	**0.97** (±.03)	**1.00** (±.00)	**1.00** (±.00)	**1.00** (±.00)
	MisOrient	$\mathbb{R}^{11}/\mathbb{R}^1/S$	**1.00** (±.00)	**1.00** (±.00)	**1.00** (±.00)	0.56 (±.10)	0.50 (±.11)	**0.99** (±.02)	**0.98** (±.02)	**0.99** (±.02)	**0.98** (±.03)	**0.99** (±.02)
	ActiveTrack	$\mathbb{R}^{10}/\mathbb{R}^3/D$	0.79 (±.08)	0.67 (±.08)	**0.95** (±.01)	0.07 (±.06)	**0.92** (±.06)	0.81 (±.05)	**0.94** (±.01)	**0.96** (±.01)	0.51 (±.12)	**0.94** (±.01)
PSM	Aggregate	–	0.0 (±.00)	0.00 (±.00)	0.40 (±.05)	0.00 (±.00)	0.02 (±.02)	0.80 (±.04)	0.00 (±.00)	0.85 (±.06)	0.46 (±.19)	**0.89** (±.03)
	NeedleReach	$\mathbb{R}^{13}/\mathbb{R}^5/S$	**1.00** (±.00)	**1.00** (±.00)	**1.00** (±.00)	0.07 (±.09)	0.89 (±.06)	**1.00** (±.00)	**0.99** (±.02)	**1.00** (±.00)	0.94 (±.20)	**1.00** (±.00)
	GauzeRetrieve	$\mathbb{R}^{25}/\mathbb{R}^5/S$	0.00 (±.00)	0.00 (±.00)	0.07 (±.05)	0.00 (±.00)	0.01 (±.02)	0.63 (±.11)	0.00 (±.00)	**0.71** (±.16)	0.43 (±.43)	**0.73** (±.12)
	NeedlePick	$\mathbb{R}^{25}/\mathbb{R}^5/S$	0.00 (±.00)	0.00 (±.00)	0.21 (±.06)	0.00 (±.00)	0.02 (±.02)	**0.91** (±.05)	0.00 (±.00)	**0.96** (±.05)	0.26 (±.33)	**0.94** (±.05)
	PegTransfer	$\mathbb{R}^{25}/\mathbb{R}^5/S$	0.00 (±.00)	0.00 (±.00)	0.56 (±.11)	0.02 (±.05)	0.05 (±.04)	0.48 (±.22)	0.00 (±.00)	0.23 (±.23)	0.31 (±.32)	**0.73** (±.20)
Bi-PSM	Aggregate	–	0.00 (±.00)	0.00 (±.00)	0.08 (±.04)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	**0.39** (±.11)
	NeedleRegrasp	$\mathbb{R}^{41}/\mathbb{R}^{10}/S$	0.00 (±.00)	0.00 (±.00)	0.09 (±.03)	0.01 (±.00)	0.01 (±.02)	0.05 (±.08)	0.00 (±.00)	0.04 (±.07)	0.00 (±.00)	**0.63** (±.19)
	BiPegTransfer	$\mathbb{R}^{41}/\mathbb{R}^{10}/S$	0.00 (±.00)	0.00 (±.00)	0.09 (±.05)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	0.00 (±.00)	0.01 (±.02)	0.00 (±.00)	**0.18** (±.14)
	Overall	–	0.46 (±.03)	0.45 (±.01)	0.68 (±.02)	0.02 (±.02)	0.24 (±.03)	0.83 (±.05)	0.48 (±.01)	0.87 (±.03)	0.58 (±.08)	**0.92** (±.02)

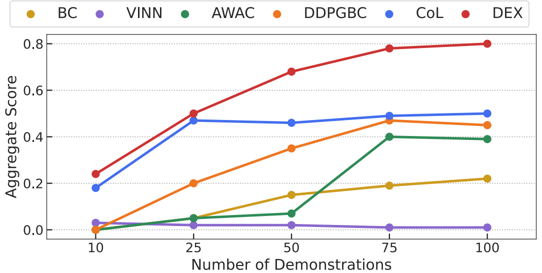
- Our method significantly outperforms prior RL-based approaches on the surgical robot learning tasks from SurRoL, especially on complex bi-manual tasks

Ablation

- Ablation on model components

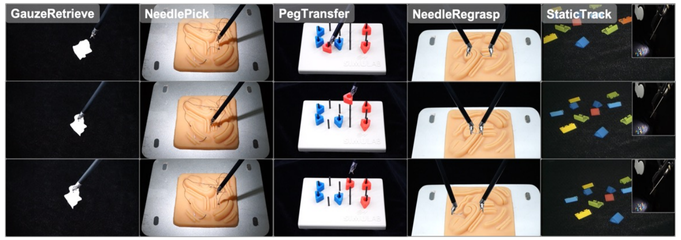


- Ablation on demonstration amount



Robot Evaluation

- Trajectories deployed on real dVRK platform



- Robot experiments on real-world tasks

Method	GauzeRetrieve	NeedlePick	PegTransfer	NeedleRegrasp	StaticTrack
BC	0.00	0.85	0.00	0.40	1.00
DDPGBC	0.75	0.95	0.35	0.65	1.00
DEX (ours)	**0.90**	**0.95**	**0.75**	**0.90**	**1.00**